DOI: 10.3217/978-3-99161-050-2-154

Optimizing Imagined Speech Decoders: A Comprehensive Study on Intra-Subject Classification

J. Muñoz-Montes de Oca [1*], A. Torres-García[1], C. Reyes-García[1], L. Villaseñor-Pineda[1]

[1] *Biosignals Processing and Medical Computing Laboratory, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico;*

*Luis Enrique Erro #1, Sta. María Tonanzintla, 72840 San Andrés Cholula, Puebla, Mexico. E-mail: jenny.munoz@inaoep.mx

Introduction: Imagined Speech (IS) refers to the mental pronunciation of words without emitting audible sounds. This task would allow the development of Brain-Computer Interfaces (BCIs) to facilitate communication for impaired people. Although deep learning (DL) algorithms have achieved promising classification rates in other neuroparadigms; their large data requirements, high computational demands, and training times limit their adoption in BCI-based IS, especially when multiclass IS tasks are involved.

Methods and Results: In this work, we analyzed the 2020 International BCI Competition dataset (Track #3) composed of 64-channel EEG signals from 15 participants imagining three words ("hello," "yes," "stop") and two phrases ("help me," "thank you") relevant for basic communication. Three preprocessing techniques were analyzed: bandpass filtering (0.1–45 Hz), detrending, common average referencing (CAR), and wavelet-based automatic tunable artifact rejection (ATAR), alongside 6 feature extraction techniques: statistical features, fractal dimension, discrete and continuous wavelet transforms (DWT, CWT), wavelet scattering transform (WST), and empirical mode decomposition-fractal (EMD-Fractal). Also, three classifiers (KNN, SVM and Random Forest (RF)) were evaluated in an intra-subject approach.

Statistical analysis identified CAR as the optimal preprocessing technique along with RF classifier. Three feature extraction techniques demonstrated comparable performance for IS classification, namely Fractal (58.0 ± 9.7%), WST (using KNN, 64.7 ± 13.0%), and EMD-Fractal (62.5 ± 11.2%) (See Fig. 1). Bayesian Optimization improved accuracy for 6 of 15 subjects, with average accuracies of 60.0 ± 10.3% (Fractal-RF) and 64.1 ± 8.4% (EMD-Fractal-RF). The proposed method achieves results similar to those reported in [1, 2], which used DL techniques with the drawbacks mentioned above. In contrast, the proposed methodology demonstrates training efficiency, with the best classifier (EMD-Fractal) requiring less than 1.5 minutes to train.

Conclusion: In this study, EMD-Fractal features which measure the complexity of EEG signals, were found to be important for IS decoding. Further, this study provides a comprehensive evaluation of different preprocessing, feature extraction, and classification techniques, resulting in a time efficient methodology for IS classification.

Acknowledgments: The first author thanks the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti) for supporting this work (scholarship CVU #1231373).

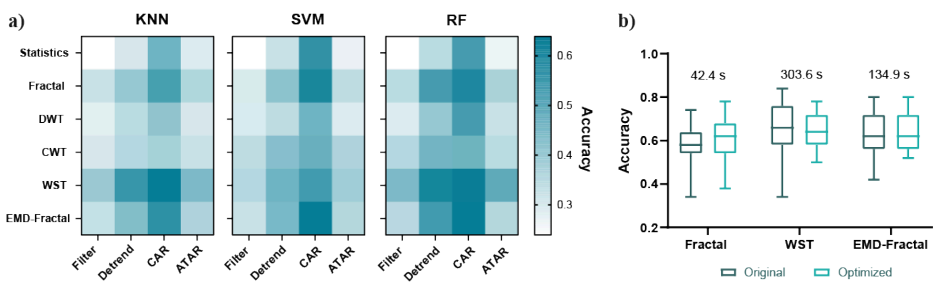

Figure 1: *a) Heatmaps show the average classification accuracy for each pipeline configuration (accuracy values are normalized from 0 to 1). b) Comparison of average accuracy for the best-performing pipelines before and after Bayesian optimization. All pipelines use CAR and RF, except WST Original, which uses KNN. Average training times (in seconds) are displayed above the boxplots.*

References:

[1] Lee, BH, Kwon, BH, Lee, DY, Jeong, JH. Speech imagery classification using length-wise training based on deep learning. In *Proceedings of the 9th International Winter Conference on Brain-Computer Interface (BCI)*, 1–5, 2021.

[2] Ko, W, Jeon, E, Suk, HI. Spectro-spatio-temporal eeg representation learning for imagined speech recognition. In *Asian Conference on Pattern Recognition*, 335–346, 2021.

Published by Verlag der Technischen Universität Graz